JSRT JOURNAL
JOURNAL OF SCIENTIFIC RESEARCH AND TECHNOLOGY
PAGES: 15-28
06/01/25
VOLUME-3 ISSUE-1 JANUARY

Analysis And Prediction Of Crime Against Women Using Machine Learning

Keerti[1], Mrs. Neha Deshmukh[2]

[1]*Student, Master of Computer Application, VTU CPGS, Kalaburagi, Karnataka, India.*
[2]*Assistant Professor, Master of Computer Application, VTU CPGS, Kalaburagi, Karnataka, India.*

ABSTRACT

A critical worldwide issue, the rising number of crimes perpetrated against women calls for fresh approaches to the problem's prevention and solution. The purpose of this research is to examine crime data from the past, find trends and patterns, and identify possible crime hotspots by using M-L algorithms. Using predictive modeling techniques and complex data analytics, this research aims to provide important information that lawmakers and law enforcement may use to protect women and reduce the likelihood that they will be victims.

Gathering and preparing massive datasets including geographical data, socioeconomic variables, criminal records, and demographic information is the meat and potatoes of this project. Publicly accessible crime reports, police records, and government databases are some of the sources of datasets. For accurate and trustworthy analysis, a comprehensive preparation procedure is carried out to handle missing numbers, normalize the data, and eliminate inconsistencies.

Keywords: Crime Prediction, Machine Learning, Predictive Modeling.

1. INTRODUCTION

This study forecasts the rates of crime against women in different states of India using a combination of random forest and linear regression. Results from linear regression outperformed those from random forest. Combating the alarming increase in violence against women in India requires accurate crime prediction. The incorporation of these techniques into preventative and real-time monitoring systems might be the subject of future studies. Data gathered from law enforcement agent interviews, questionnaires, and document analysis forms the basis of an intelligence and crime analysis model. For efficient analysis of criminal data, the model—created in Excel and represented in UML—integrates big data analytics, M-L, and A-I. While the results demonstrate a considerable advance in the efficiency and accuracy of crime analysis, they also demonstrate the predominance of conventional methods. For the purpose of predicting both high-crime regions and age groups prone to criminal behavior, the suggested approach employs an optimized K-means algorithm. Clustering comparable data points allows this approach to enhance efficiency and decrease time complexity in crime pattern prediction. If you want more precise predictions, use the elbow approach to find out how many clusters are best. The study shows that K-means is a good tool for pinpointing areas with high crime rates and vulnerable populations. Improving crime pattern analysis might be the subject of future research that aims to refine the algorithm, include other data sources, and investigate more sophisticated machine learning methods.
For the purpose of evaluating crime data pertaining to crimes against women, the research assesses six machine learning algorithms: KNN, decision trees, Naïve Bayes, Linear Regression, CART, and SVM. The predicted accuracy of each algorithm is evaluated using methods like k-fold cross-validation. This study demonstrates how these algorithms might help with crime prevention and enforcement by reliably forecasting future crime rates. To further improve the accuracy of predictions and the efficacy of models, future studies may use more algorithms, larger datasets, and more current data sets. This research uses SPSS 23 for statistical analysis and a structured questionnaire survey to look at the effects of crimes against women in Delhi's NCT. Findings show a correlation between environmental factors and crime rates, with low-crime regions enjoying sufficient amenities and high-crime areas lacking in attractive ones. To further understand the factors that influence crime rates, future studies should include more variables and use larger samples.

1.1.1 OBJECTIVES

In order to accurately assess and forecast crimes against women, this research aims to construct an M-L model that can detect trends and patterns in crime data. In order to anticipate crime hotspots and at-risk populations, the

project intends to use and assess several algorithms, such as K-NN, Decision Trees, Naïve Bayes, and SVM. With the use of data preprocessing and feature engineering, the study also hopes to improve the model's prediction accuracy, which will provide useful feedback for future work.

1.1.2 METHODOLOGY

- **Data Collection:** Collect all available datasets on crimes against women, including incident reports, demographic information, geographic locations, and time stamps. Police reports, public documents, and databases are among possible sources.
- **Data Preprocessing:** Remove errors, outliers, and missing values from the acquired data. Make sure the data is compatible with machine learning algorithms by normalizing and standardizing it. Develop useful predictors by extracting features and engineering them.
- **Conducting EDA (Exploratory Data Analysis):** Apply EDA to comprehend the data distribution, spot important patterns, and see trends graphically. This stage is useful for choosing the right characteristics.
- **Algorithm Selection and Implementation:** For model creation, choose machine learning algorithms like Support Vector Machines (SVM), Decision Trees, Naïve Bayes, K-Nearest Neighbors (KNN), Linear Regression, CART, and Naïve Bayes. Put these algorithms into action by selecting an appropriate programming language and ML framework.
- **Model Training and Validation:** Separate the dataset and test subsets. Utilize techniques such as k-fold cross-validation to train the models and verify their evaluate performance. Adjust the model's parameters to your liking and determine its accuracy, precision, F1 score, and recall.
- **Analysis and Prediction:** Make use of the trained models to pinpoint areas prone to crime and pinpoint populations who are particularly at risk. Use the results of the analysis to inform decisions about allocating resources and preventing crime.
- **Review and Comparison:** Evaluate several algorithms by comparing their performance in terms of accuracy and effectiveness. Based on assessment criteria, choose the model that performs the best and put it into action.
- **Application and Integration:** Create an intuitive interface or tool to connect the prediction model to data systems that process information in real-time. Because of this, lawmakers and police officers may use the model for better decision-making and preventative measures against crime.
- **Future Work and Refinements:** Find areas that might need some work, such looking at more complex M-L approaches or adding more data sources. Keep improving the model to make it more accurate and useful.
- **Choosing the Algorithms:** Prior to making a final decision, many approaches are tested to assess their ability to forecast criminal behavior. Methods for dealing with complicated, high-dimensional data, such as Support Vector Machines (SVM) and K-Nearest Neighbors (KNN). Naïve Bayes offers a probabilistic technique for classification jobs, whereas Decision Trees and CART give interpretability and simple decision-making procedures. Predicting numerical crime rates is done using Linear Regression. By integrating various algorithms, we can conduct a thorough evaluation, compare their results, and choose the best model for crime prediction.

2. ITERATURE SURVEY

Bachchu Lal Rai and Dr. Megha Kamble [1] have investigated the decline in India's crime rate. This research, which was published in the Journal of Advances and Scholarly Researches in Allied Education, aims to predict particular sorts of crimes in certain places by building a model using training datasets that are clean and processed. To learn what influences crime rates and how to anticipate them, the study uses clustering algorithms, optimization approaches, and statistical analysis.

Adesola Falade and Ambrose Azeta [2] have performed an exhaustive literature search on data mining methods and crime prediction. The researchers used a three-pronged approach in their investigation, which included collecting appropriate literature, developing a taxonomy, and writing up the results. The paper emphasizes that supervised learning methods are widely used for crime prediction and control.

Priyanka Das and Asit Kumar Das [3] have presented a strategy for examining female-targeted crime stories published in online newspapers. Their method begins with trend analysis, then moves on to preprocessing data

extracted from news stories pertaining to crimes. Using the states as nodes, the research finds crime patterns using the Info map clustering technique, which reduces dimensions and detects communities.

Sunil et.al [4] presented an approach for detecting, analyzing, and predicting criminal patterns at the 2017 International Conference on Electronics, Communication, and Aerospace Technology (ICECA). The method improves the prediction power of crime data by using supervised, semi-supervised, and unsupervised learning strategies.

Arushi Jain et.al [5] presented a framework at the International Conference on Information Security and Privacy (ICISP2015) that used Pig in conjunction with Hadoop. To import structured data, their method makes use of Apache Sqoop, and to deal with unstructured data, they use Apache Flume.

3. TOOLS AND TECHNOLOGIES USED

3.1 Python Implementation: Scripts and Programs

Machine learning analysis and prediction of crimes against women may be accomplished well using Python. When it comes to automating activities and executing sophisticated algorithms without human input, scripts are crucial in Python. For effective model testing and improvement, these scripts may be run repeatedly and include Python code. Improving crime prediction models relies heavily on this iterative method.

3.2 Key Python Libraries for Crime Prediction

NumPy
Pandas
Scikit-Learn
Tensor Flow and Keas
Matplotlib and Seaborn
Stats models

3.3 Utilizing Python for Web Frameworks and Database Interactions

Web framework and database integration is crucial for an all-inclusive criminal analysis system:

- **Flask and Django:** These popular Python web frameworks enable the building of online applications that deploy crime prediction algorithms. While Django is more organized and comes with built-in tools for security and database administration, Flask is more open and simple.
- **SQLite:** SQLite is a handy tool for lightweight, file-based databases. The ability to store and retrieve data efficiently is crucial for crime data management, and it provides that.
- **MySQL:** Perfect for bigger datasets used in criminal investigation, MySQL provides strong relational database administration and wide support for data operations, making it an excellent choice for more complex database needs.

3.3.1 Testing and Debugging in Python

The development of trustworthy M.L. models relies heavily on testing. Building test cases and ensuring code functioning is made easier using Python's built-in modules like pytest and unit test.

3.3.2 Functions and Modular Programming

Python functions make code more readable and easier to maintain by separating it into reusable pieces. With modular programming, you may build reusable, testable components that can be utilized in several applications.

Developers may create a model that predicts crimes against women by using Python's many modules, frameworks, and tools. Python is an excellent option for handling complicated data analysis and M-L jobs due to its adaptability and its vast ecosystem of tools.

3.4 Algorithm- SVM Algorithm

Overview of SVM:

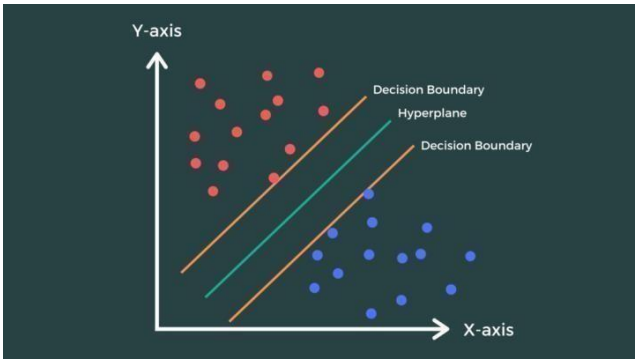

Fig:1 SVM Overview

4. HARDWARE AND SOFTWARE REQUIREMENTS

4.1 HARDWARE REQUIREMENTS

TABLE-1

1	Processor	Intel Core i3 or above
2	Processor Speed	2.10 GHz
3	RAM	4GB or above
4	Hard Disk	256GB SSD or 500 GB HHD
5	Monitor	16.5 inch
6	Keyboard	Standard keyboard QWERTY (108keys)
7	Mouse	Option Mouse

4.2 SOFTWARE REQUIREMENTS

TABLE-2

1	OS	Windows 10 or higher
2	Backend Programming Language	Python
3	Frontend Programming Languages	HTML, CSS, JavaScript
4	Web Framework	Flask or Django (for Python)
5	IDE	PyCharm, Visual Studio Code, or Jupyter Notebook
6	Dataset	Public datasets like the UCI Crime and Communities dataset, Kaggle datasets on crime, or local government-provided crime datasets

5. SYSTEM DESIGN

5.1 SYSTEM PERSPECTIVE

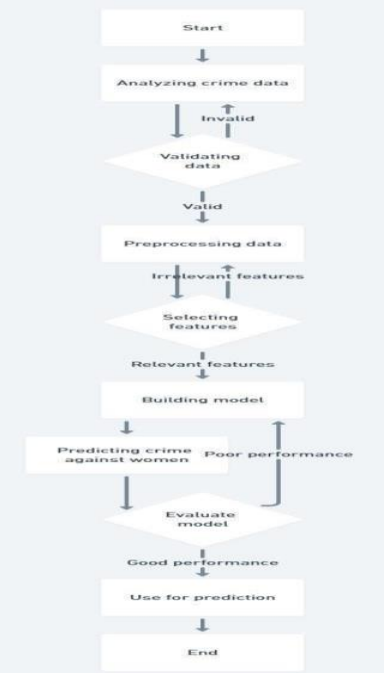

Figure .4 System perspective diagram

The system perspective diagram, shown in Figure 4, shows the overall architecture and how the many components interact with each other. The data sources shown in the graphic include criminal databases and backend data that is reported in real-time. The Python-based backend is responsible for storing data, processing it, and running machine learning models. Django is one example of a web framework that handles backend to frontend communication. Visualization of crime data and forecasts is made possible with an HTML, CSS, and JavaScript user interface.

5.2 Context Diagram

Figure .4 Context Diagram

An overarching view of the project's system's interactions with external entities is given by the context diagram.

Public crime databases, social media, and law enforcement are examples of data sources; scholars, policymakers, and the public at large are examples of end-users. Through the use of M-L algorithms, the system takes in data from a variety of sources, analyzes it, and then produces forecasts and insights on crime patterns. Through the use of real-time crime reporting and prediction warnings, law enforcement authorities may take preemptive steps.

5.2 DETAILED DESIGN

5.2.1 USECASE DIAGRAM

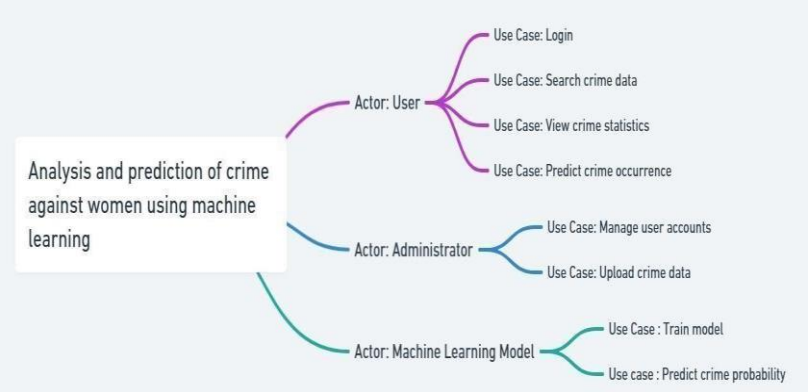

Fig-5 Use-case for users interacts with Structure & designates

The use case diagram describes the main features and how users interact with the system. Everyone from experts to members of the general public and law enforcement are key players. The primary use cases are data input, where users may submit crime reports and associated data; data analysis, where the system applies machine learning algorithms to process and analyze the data; and prediction, where probable future crime hotspots are anticipated. Interactive dashboards allow users to see crime patterns and produce reports for further action. You can see all the important interactions in this figure.

5.2.2 SEQUENCE DIAGRAM

Fig-6 Sequence Diagram

The sequence diagram shows the project's dynamics that will get us to our analysis and prediction goals. It all

starts with the end-user, who may be a member of the police force or a member of parliament, making a request for crime analysis or prediction via the interface. Following this, the data processing module gets the necessary information from several interconnected data sources, such as social networking networks and public criminal databases.

5.2.3 COLLABORATION DIAGRAM

Fig-6 Collaboration Diagram

In the project the collaboration diagram, depicts the structural organization between its components to achieve crime analysis and prediction. The key components include the user interface, data retrieval module, data preprocessing module, machine learning engine, and report generation module. When an end-user, such as a law enforcement officer or policymaker, initiates a request via the user interface, the data retrieval module is triggered such as crime databases and social media platforms.

5.2.4 ACTIVITY DIAGRAM

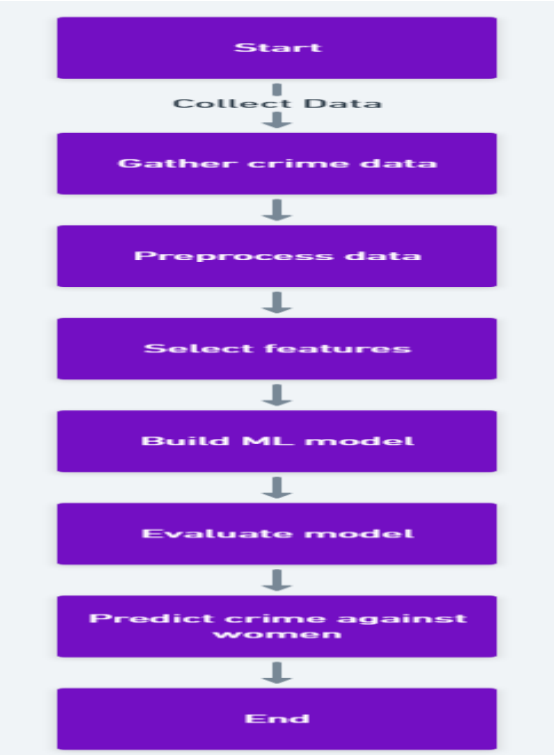

Figure.6 Activity Diagram

In order to examine and forecast crimes perpetrated against women, the system goes through a series of steps, as

shown in the activity diagram. Law enforcement officers or legislators, who are examples of end-users, begin the process by requesting analysis or prediction via the user interface, as shown in the figure. Public criminal databases and social media platforms are activated as a result. Data cleansing, normalization, and transformation into an analytically usable format are all steps in the preprocessing pipeline that follows data retrieval.

6. IMPLEMENTATION

The methodology, tools, and processes for building the system to analyze and forecast crimes perpetrated against women using a Support Vector Machine (SVM) algorithm are laid out in the Implementation portion of the Software Requirements Specification (SRS) paper.

6.1 Development Environment

This criminal prediction system's development environment comprises:

- **Programming Languages:** Python's strong support for data analysis and machine learning packages led to its selection as the main programming language.
- **Software Development Environments (IDEs):** PyCharm and Visual Studio Code are among the IDEs used by the development team. Git is used for version control, while GitHub is used to host repositories. The team communicates and keeps tabs on projects using collaboration platforms like Slack and JIRA.

6.2 System Components

Key components of the system are implemented as follows:

- **Module for Preprocessing Data:** In this section, we clean, normalize, and extract features from the data. Prior to SVM modeling, data is prepared by implementing algorithms that deal with missing values, scale numerical features, and encode categorical variables.
- **Machine Learning Model Module**: SVM algorithms form the backbone of the system. To categorize and forecast crime trends, the SVM model is built and trained with the use of crime data from the past. To get the best possible results, we validate our model and fine-tune its hyperparameters.
- **User Interface:** Analysts and law enforcement will find the user interface easy to utilize. It provides a user-friendly interface with reports and visualizations that let users enter crime-related data and examine prediction results.

6.3 Integration

Integration efforts include:

- **API Integration:** Securely transmit data to sources like social media platforms and public criminal databases via the use of APIs. The secure and uninterrupted transfer of data into the system is guaranteed by these APIs.
- **Database Integration:** In order to store preprocessed data, trained SVM models, and prediction outputs, the system integrates with databases such as SQL or NoSQL. Efficient data management and retrieval are made possible by this connection.

6.4 Deployment

Deployment details include:

- **Environment of Deployment:** Based on the organization's requirements and infrastructure, it may be implemented either on-premises or in the cloud, such as AWS or Azure.
- **Efficient Configuration Management:** All system settings, dependencies, and environment variables are carefully monitored to guarantee consistent deployment in different environments. This promotes seamless operation and scalability.

6.5 SCREENSHOTS

HOME SCREEN

Fig-7 Home Screen

SELECT DATASET

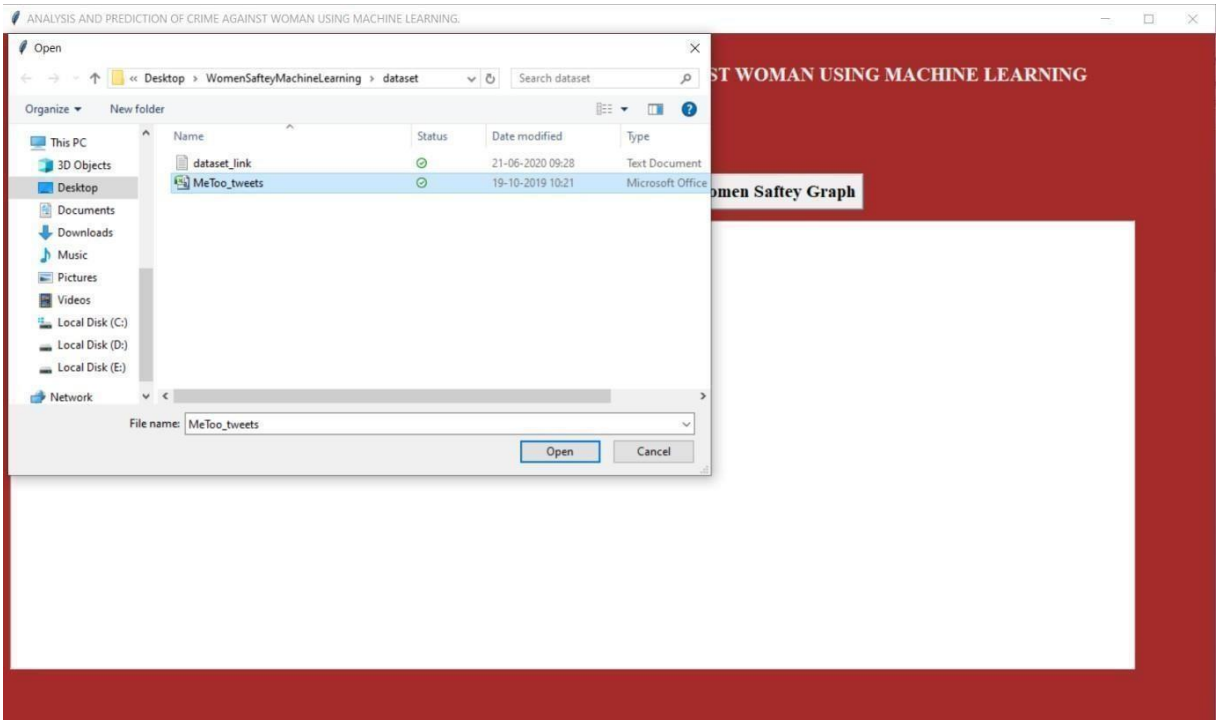

Fig-8 Select Dataset

UPLOAD DATASET

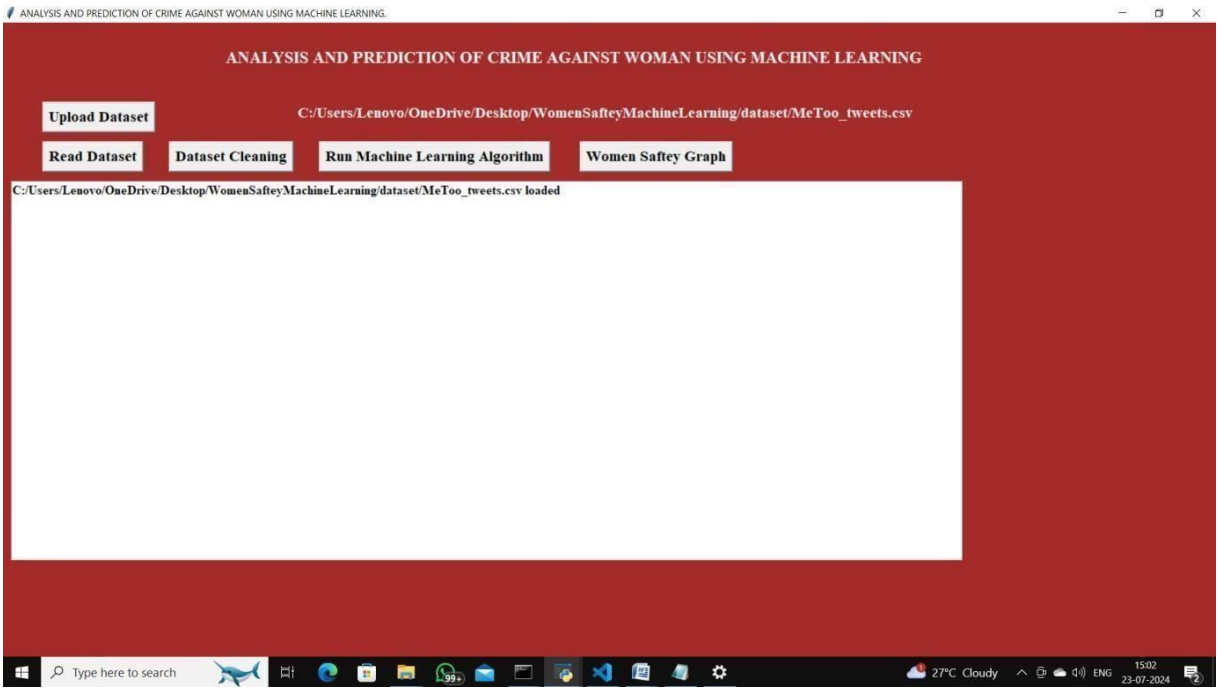

Fig-9 Upload Dataset

READ TWEETS:

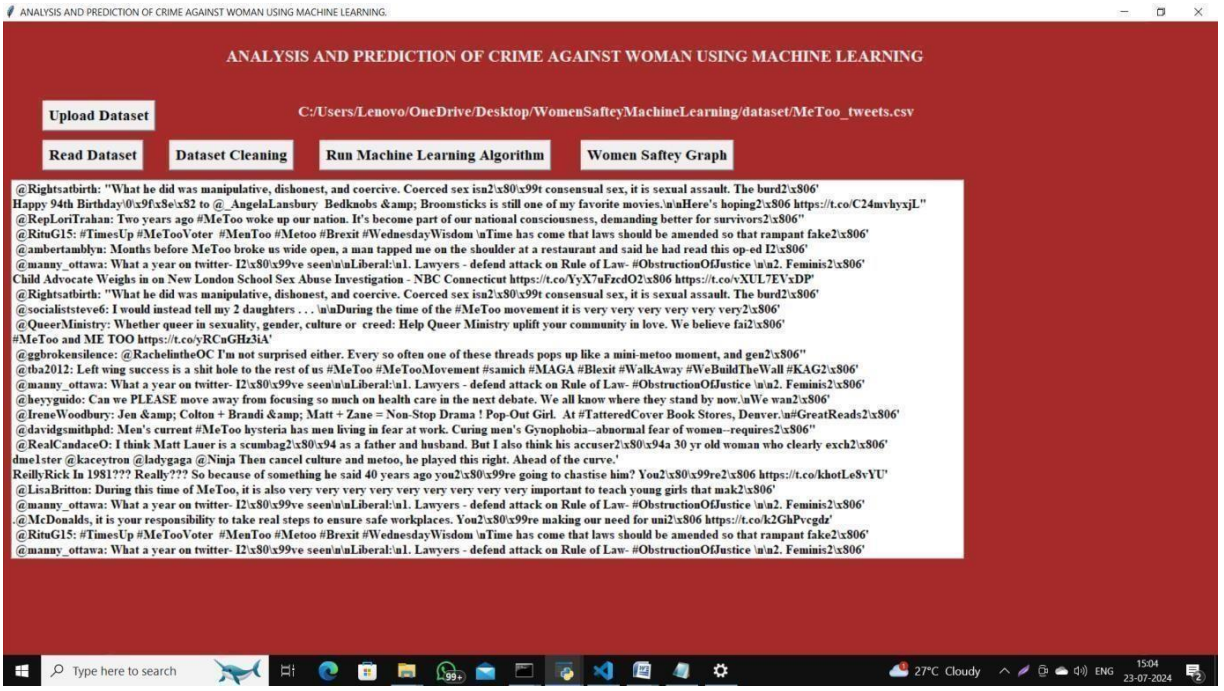

Fig-10 Read Tweets

RUN MACHINELEARNING ALGORITHM

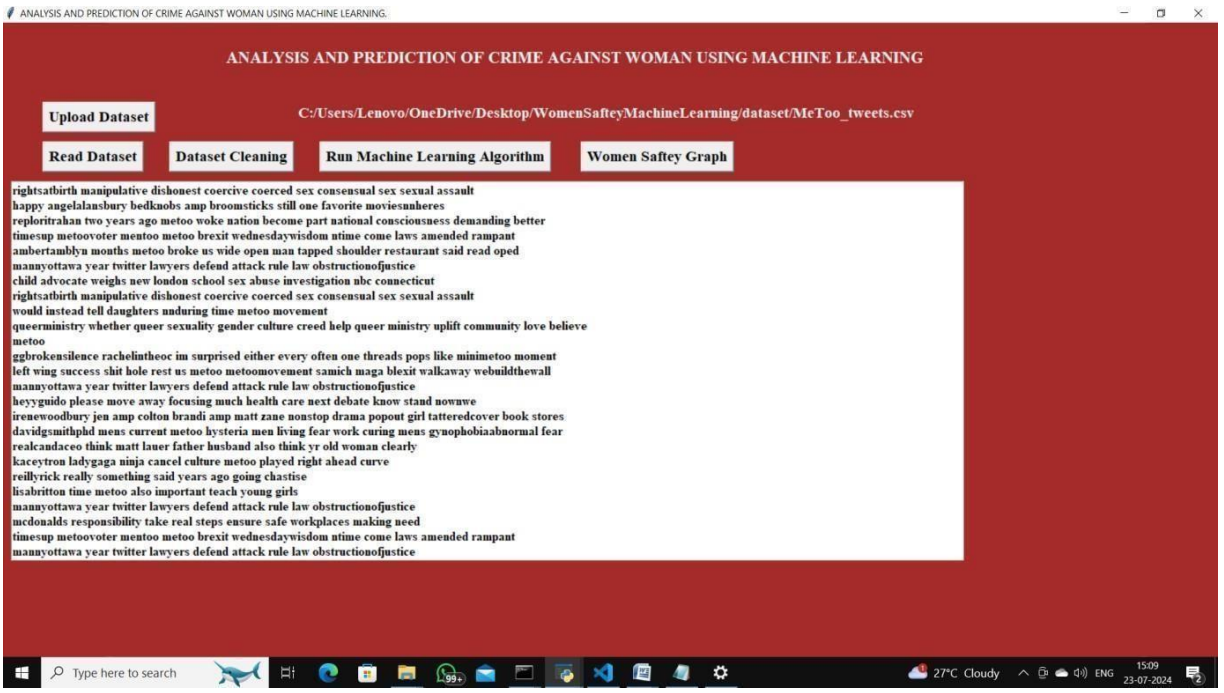

Fig-11 Run Machine Learning Algorithm

WOMEN SAFETY AND SENTIMENT GRAPH

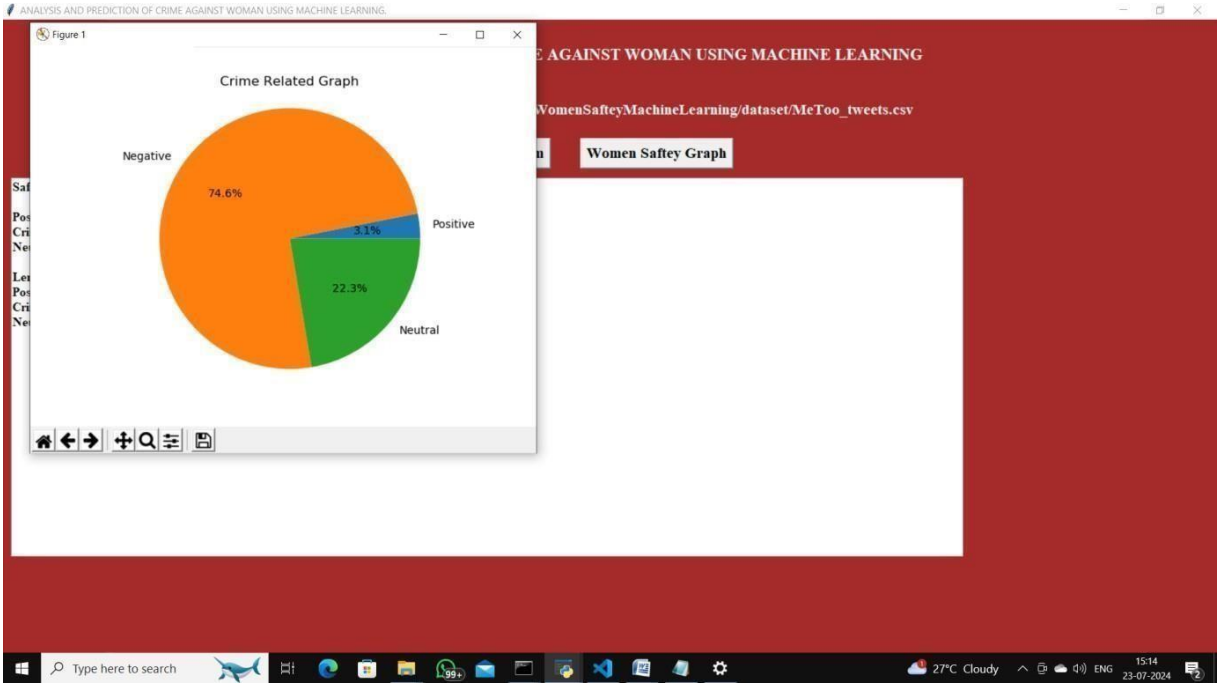

Fig-12 Women safety and Sentiment Graph

6.6 TEST CASES

TABLE - 3

Test Case ID	Test Case Description	Expected Result	Actual Result	Remarks
TC01	Upload dataset of crime reports	Dataset is uploaded successfully	Dataset can be seen uploaded	Pass
TC02	Upload unsupported file types	System shouldn't recognize unsupported files	System doesn't recognize unsupported file	Pass

7 TEST RESULTS

A successful validation of the system's functions is shown by the test results for the project. Uploading the dataset of crime reports was a success in Test Case TC01, proving that the system could effectively process data input. Data integrity was ensured in Test Case TC02, which required the system to analyze the dataset for missing values. It successfully found these gaps.

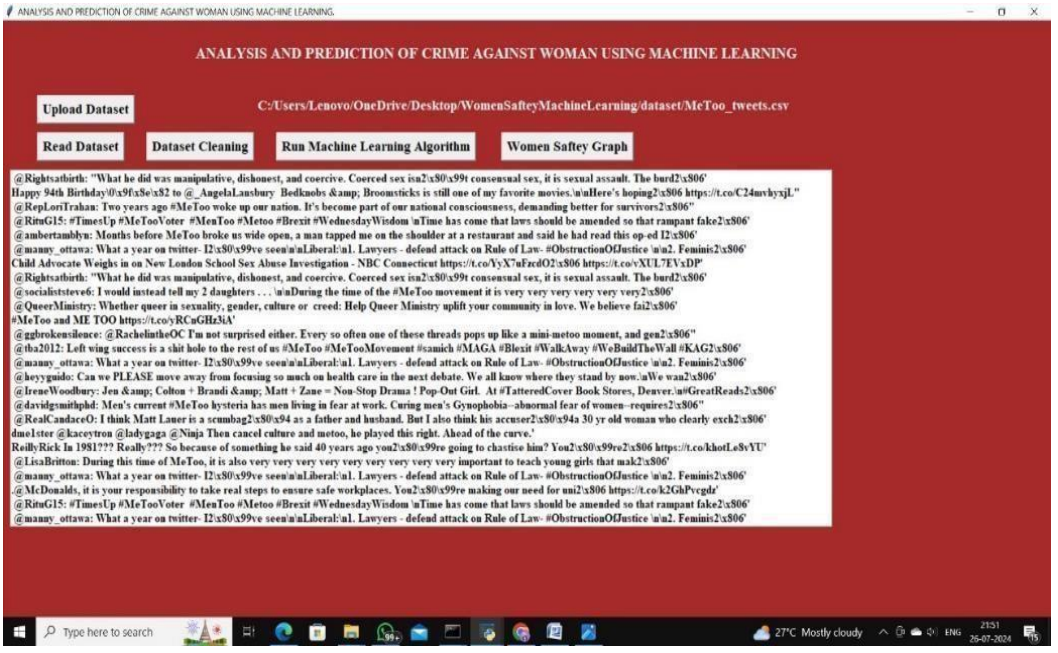

Fig-13 Dataset is uploaded successfully

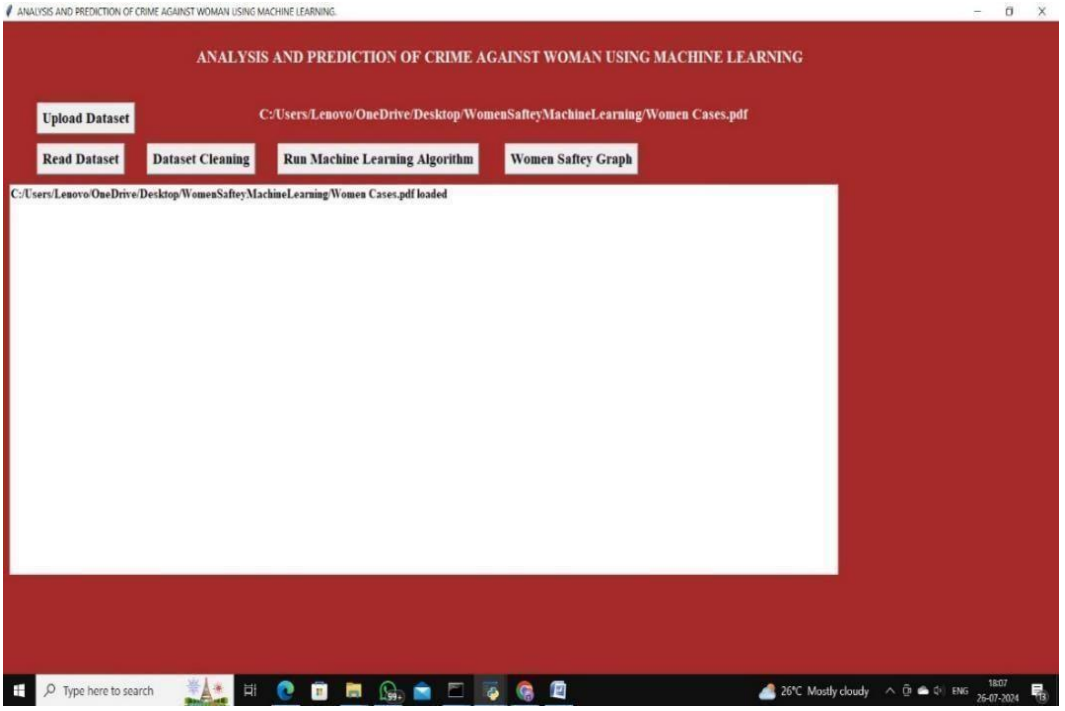

Fig-14 Unsupported file doesn't get recognized

8 CONCLUSION

By completing the project, we have shown that cutting-edge technology can solve important social problems. Law enforcement and lawmakers may benefit from this system's data-driven approach to crime pattern prediction and analysis, made possible by using M-L algorithms. Both the identification of high-risk regions and the formulation of efficient crime prevention methods are aided by the complete study. This study highlights how machine learning can empower communities and enhance public safety via the integration of varied data sources and powerful prediction models. With further development and implementation, this system has the potential to greatly aid in the decrease of crimes and violence against women, guaranteeing a society free from such threats.

9 FUTURE ENHANCEMENT

Its efficacy and influence might be further enhanced with a number of planned upgrades. Predictions may be much improved in terms of accuracy and timeliness if many sources of real-time data were integrated, for example, public reporting, emergency response systems, and social media. We may learn more about the dynamics of crime if we broaden our focus to include socioeconomic status, indications of mental health, and environmental variables, among others. More actionable information and faster decision-making might be possible with the development of user-friendly dashboards and interfaces for lawmakers and police officers. Maintaining the system's relevance and effectiveness in responding to changing crime trends requires ongoing model improvement and validation using new data. These enhancements have the potential to promote more effective techniques for crime prevention and intervention while also improving the system's overall efficacy.

BIBLIOGRAPHY

1. Abubakar Aliyu Machina, Li Songjiang, "Crime Analysis and Intelligence System Model Design Using Big Data," *International Journal of Computer Applications* (0975 – 8887), Volume 175– No. 22, October 2020.
2. Krishnendu S.G, Lakshmi P.P, Nitha L, "Crime Analysis and Prediction Using Optimized K-Means Algorithm," *IEEE Xplore*, Part Number: CFP20K25-ART; ISBN: 978-1-7281-4889-2, 2020.
3. P. Tamilarasi, Dr. R.Uma Rani, "Diagnosis of Crime Rate Against Women Using K- Fold Cross Validation Through Machine Learning Algorithms," *IEEE Xplore*, Part Number: CFP20K25-ART; ISBN: 978-1-7281-4889-2, 2020.
4. Charu Nangia, Dr. Devendra Pratap Singh, Dr. Sabir Ali, "Built Environment and Its Impact on CrimesRelated to Women in NCT of Delhi: A Pilot Survey," *ISSN Print: 0976-6480 and ISSN Online: 0976-6499*, 2019.
5. Bachchu Lal Rai, Dr. Megha Kamble, "Design and Analysis of Machine Learning Algorithms for the Reduction of Crime Rates in India," *Journal of Advances and Scholarly Researches in Allied Education*, Vol. 16, Issue No. 11, November 2019, ISSN 2230-7540.
6. Adesola Falade, Ambrose Azeta, "Systematic Literature Review of Crime Prediction and Data Mining," *Review of Computer Engineering Studies*, Vol. 6, No. 3, September 2019, pp. 56-63, Journal Homepage.
7. Priyanka Das, Asit Kumar Das, "Crime Analysis Against Women from Online Newspaper Reports and an Approach to Apply It in Dynamic Environment."
8. Sunil Yadav, Meet Timbadi, Ajit Yadav, Rohit Vishwakarma, "Crime Pattern Detection, Analysis & Prediction," *International Conference on Electronics, Communication and Aerospace Technology (ICECA 2017)*, 2017.
9. Arushi Jain, Vishal Bhatnagar, "Crime Data Analysis Using Pig with Hadoop," *International Conference on Information Security & Privacy (ICISP 2015)*, December 11-12, 2015, Nagpur, INDIA.
10. Vineet Pande, Viraj Samant, Sindhu Nair, "Crime Detection Using Data Mining," *International Journal of Engineering Research & Technology (IJERT)*, ISSN: 2278- 0181, Vol. 5, Issue 2016. *Procedia Computer Science*, 55 (2015), pp. 251 – 260.
11. Chih-Hao Ku, Gondy Leroy, "A Decision Support System: Automated Crime Report Analysis and Classification for E-Government," *Journal*, DOI:10.1016/j.giq.2014.08.003, ISSN: 0740-624X, © 2014 Elsevier Inc.
12. Mohammad A., Patricia L. Brantingham, "CRIMETRACER: Activity Space Based Crime Location Prediction," *IEEE/ACM International Conference on Advances inSocial Networks Analysis and Mining (ASONAM 2014)*, 2014.
13. [14]. Xingan Li, Martti Juhola, "Country Crime Analysis Using the Self-Organizing Map, with Specia.
14. Regard to Demographic Factors," *Springer-Verlag London*, DOI:10.1007/s00146-013-0441-7, 2013.
15. Priyanka Kulkarni, & Dr. Swaroopa Shastri. (2024). Rice Leaf Diseases Detection Using Machine Learning. Journal of Scientific Research and Technology, 2(1), 17–22. https://doi.org/10.61808/jsrt81
16. Shilpa Patil. (2023). Security for Electronic Health Record Based on Attribute using Block-Chain Technology. Journal of Scientific Research and Technology, 1(6), 145–155. https://doi.org/10.5281/zenodo.8330325
17. Mohammed Maaz, Md Akif Ahmed, Md Maqsood, & Dr Shridevi Soma. (2023). Development Of Service Deployment Models In Private Cloud. Journal of Scientific Research and Technology, 1(9), 1–12. https://doi.org/10.61808/jsrt74
18. Antariksh Sharma, Prof. Vibhakar Mansotra, & Kuljeet Singh. (2023). Detection of Mirai Botnet Attacks on IoT devices Using Deep Learning. Journal of Scientific Research and Technology, 1(6), 174–187.
19. Dr. Megha Rani Raigonda, & Shweta. (2024). Signature Verification System Using SSIM In Image Processing. Journal of Scientific Research and Technology, 2(1), 5–11. https://doi.org/10.61808/jsrt79
20. Shri Udayshankar B, Veeraj R Singh, Sampras P, & Aryan Dhage. (2023). Fake Job Post Prediction Using Data Mining. Journal of Scientific Research and Technology, 1(2), 39–47.
21. Gaurav Prajapati, Avinash, Lav Kumar, & Smt. Rekha S Patil. (2023). Road Accident Prediction Using Machine Learning. Journal of Scientific Research and Technology, 1(2), 48–59.
22. Dr. Rekha Patil, Vidya Kumar Katrabad, Mahantappa, & Sunil Kumar. (2023). Image Classification Using CNN Model Based on Deep Learning. Journal of Scientific Research and Technology, 1(2), 60–71.
23. Ambresh Bhadrashetty, & Surekha Patil. (2024). Movie Success and Rating Prediction Using Data Mining. Journal of Scientific Research and Technology, 2(1), 1–4. https://doi.org/10.61808/jsrt78
24. Dr. Megha Rani Raigonda, & Shweta. (2024). Signature Verification System Using SSIM In Image Processing. Journal of Scientific Research and Technology, 2(1), 5–11. https://doi.org/10.61808/jsrt79